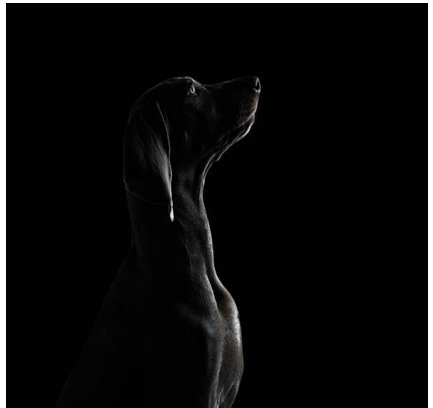
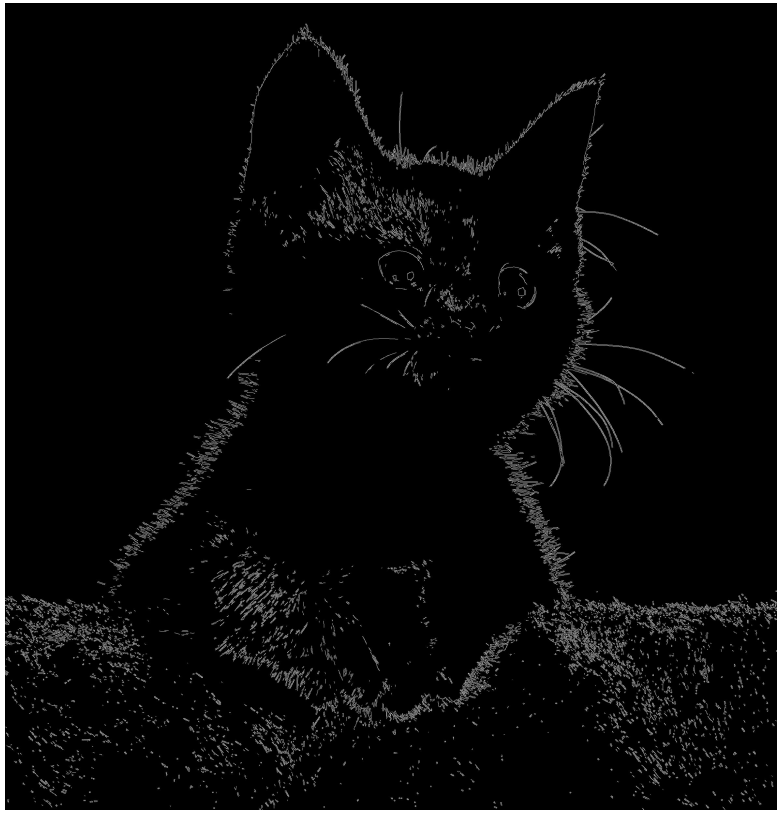
Кошечки или собачки





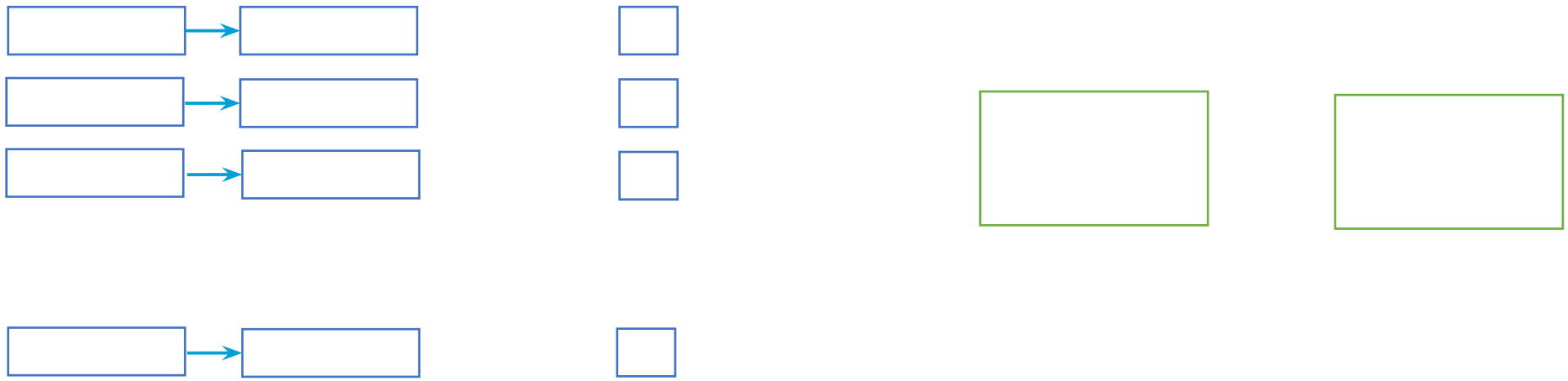
Алгоритм?



Машинное обучение
Machine Learning

Machine learning is a field of computer science that gives computers the ability to learn without being explicitly programmed.[1]

Wikipedia



Street View House Numbers

32x32
pixels
10 classes
~70000
train
~25000
test

http://ufldl.stanford.edu/housenumbers

Метод ближайших соседей
Nearest neighbor

```
Train:
    просто все запомнить
```

```
Predict:
    найти ближайший  и
выдать его класс
```

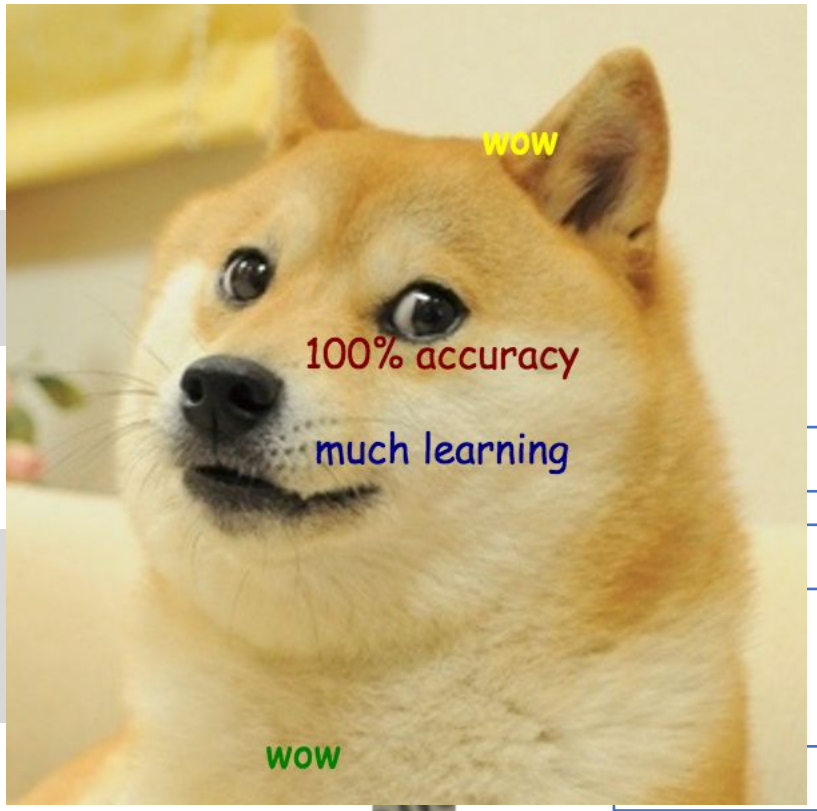
$$L_2 = \sqrt{\sum_i (v_i - u_i)^2}$$

$$L_1 = \sum_i |v_i - u_i|$$

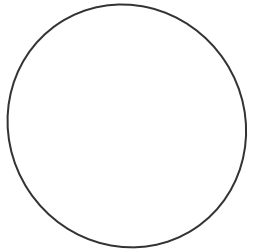
Точность на тренировочных vs тестовых данных

```
Train:
    просто все запомнить
```

```
Predict:
    найти ближайший  и
выдать его класс
```

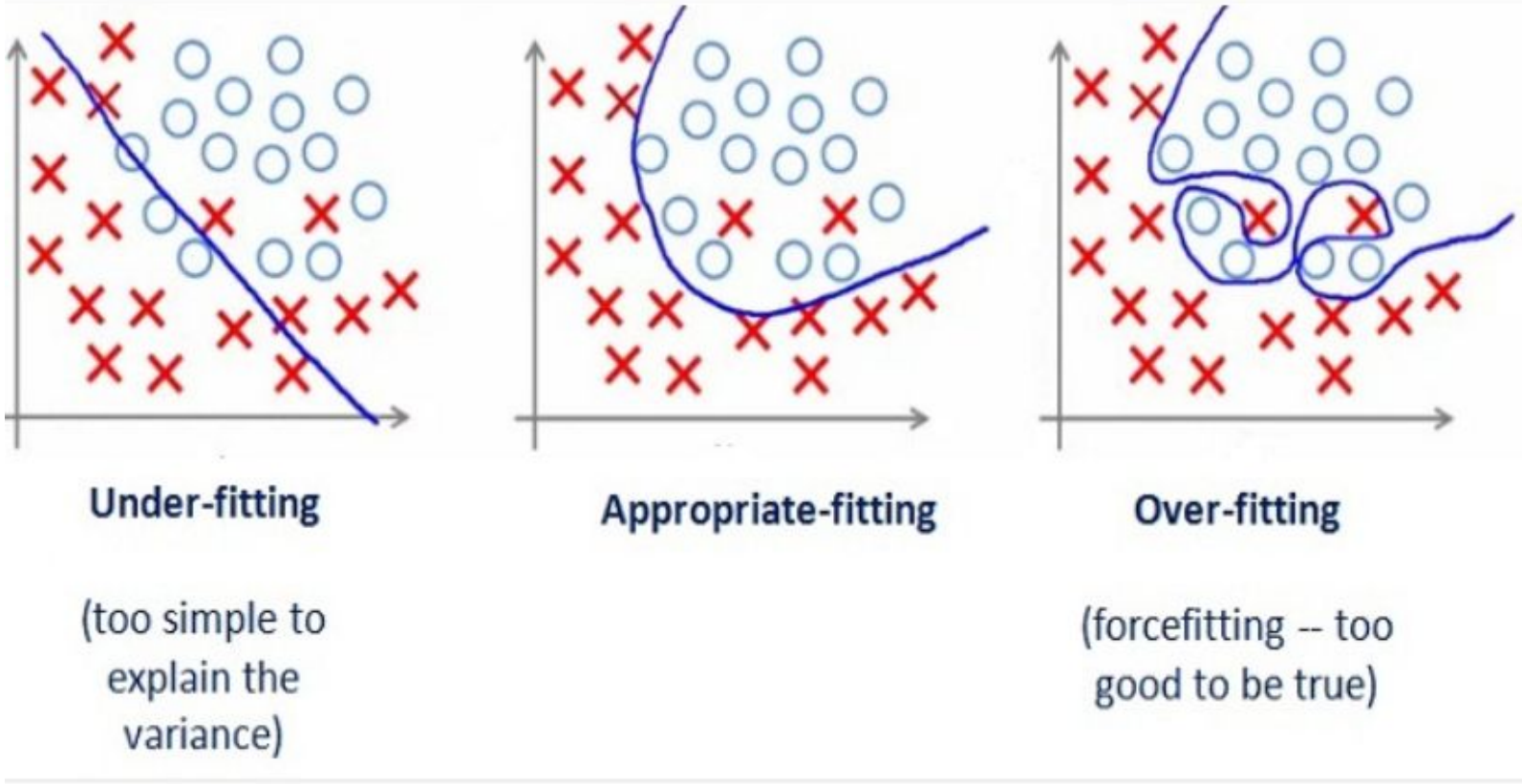


Метод k-ближайших соседей
K-nearest neighbors

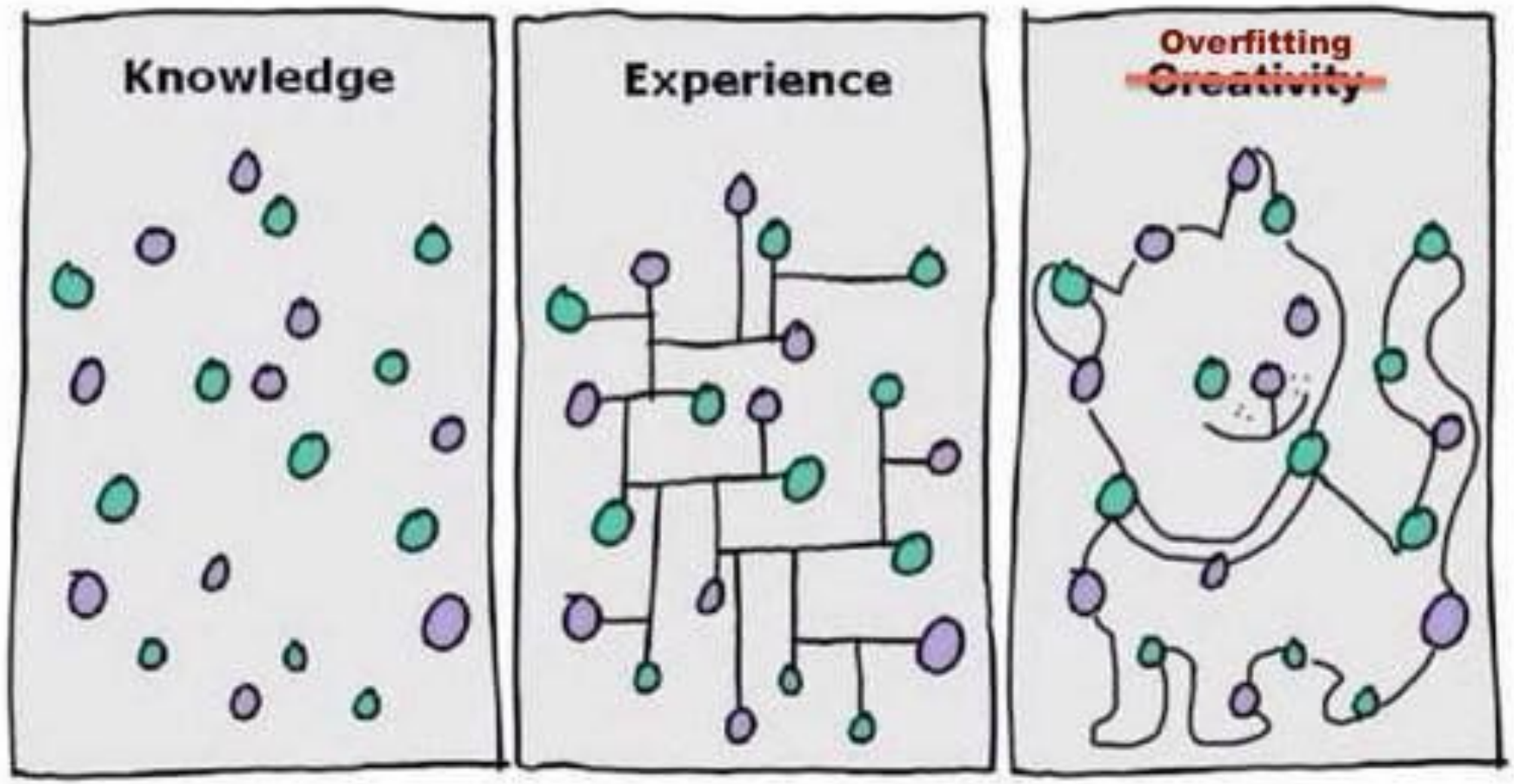


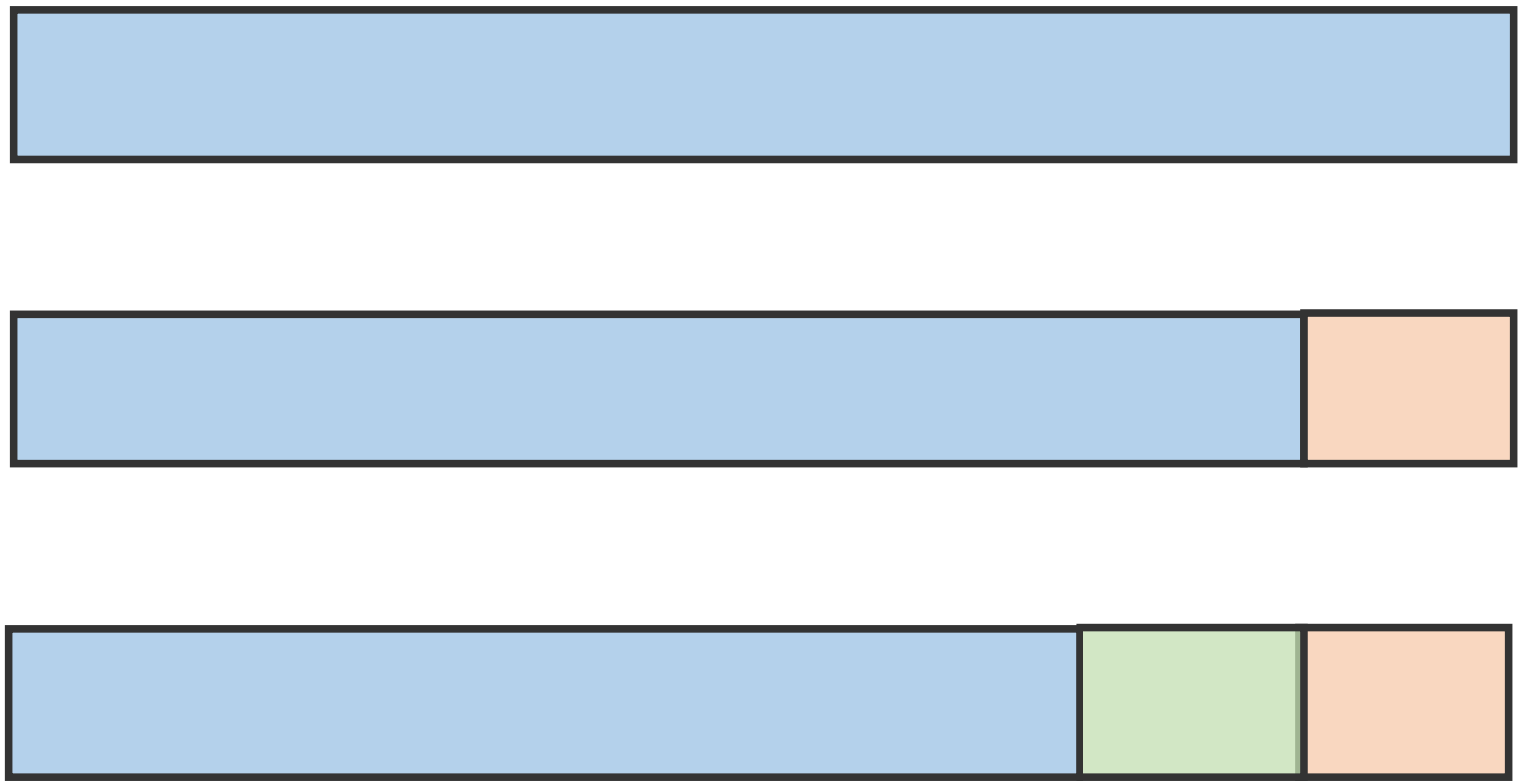
Как выбрать K?

Переобучение и недообучение
Overfitting vs underfitting

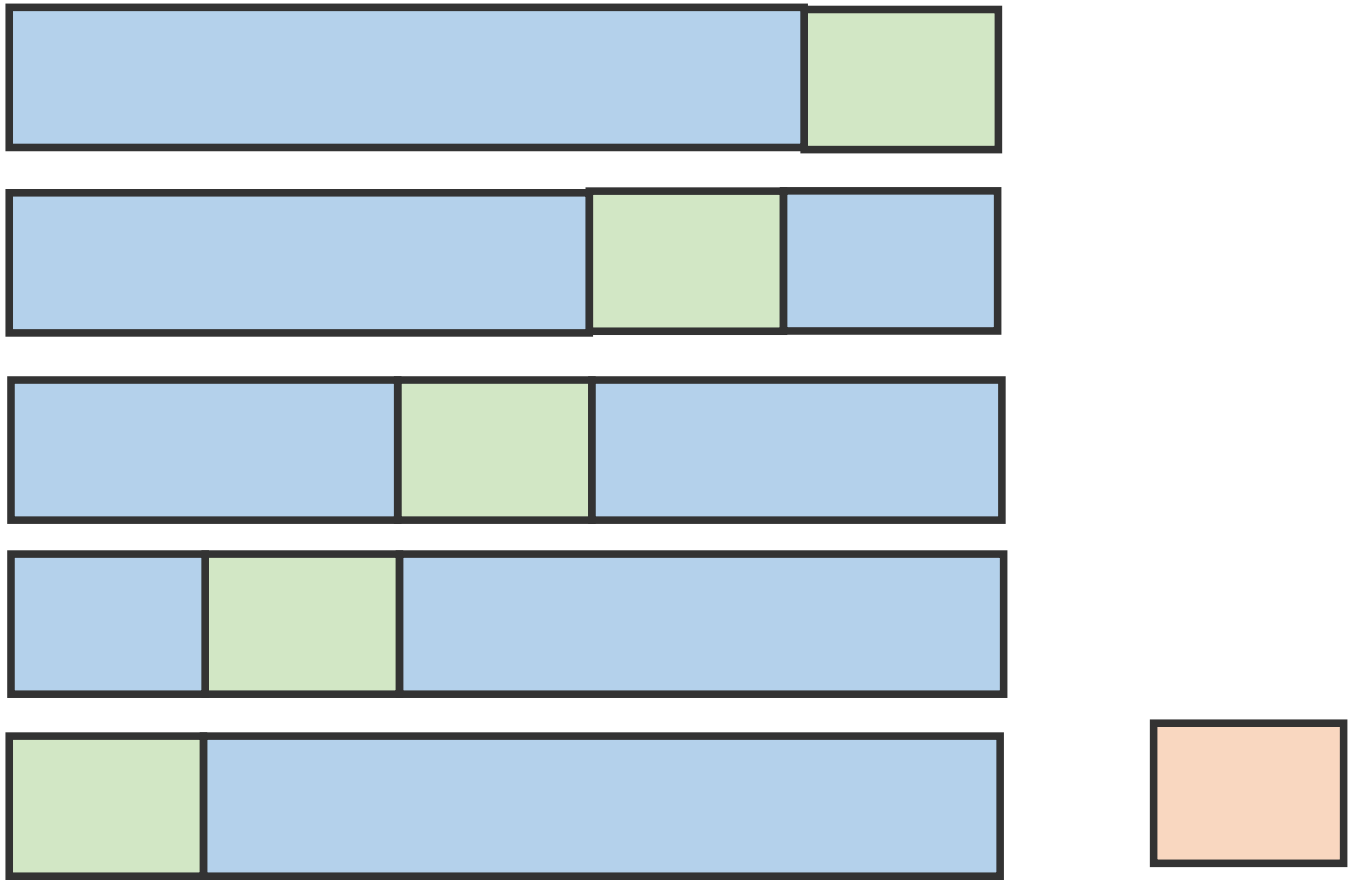


What is underfitting and overfitting in machine learning and how to deal


Knowledge
Experience
Overfitting
Creativity



Кросс-валидация
Cross-validation



Как сравнивать?

Бинарная классификация
Binary classification

v
s

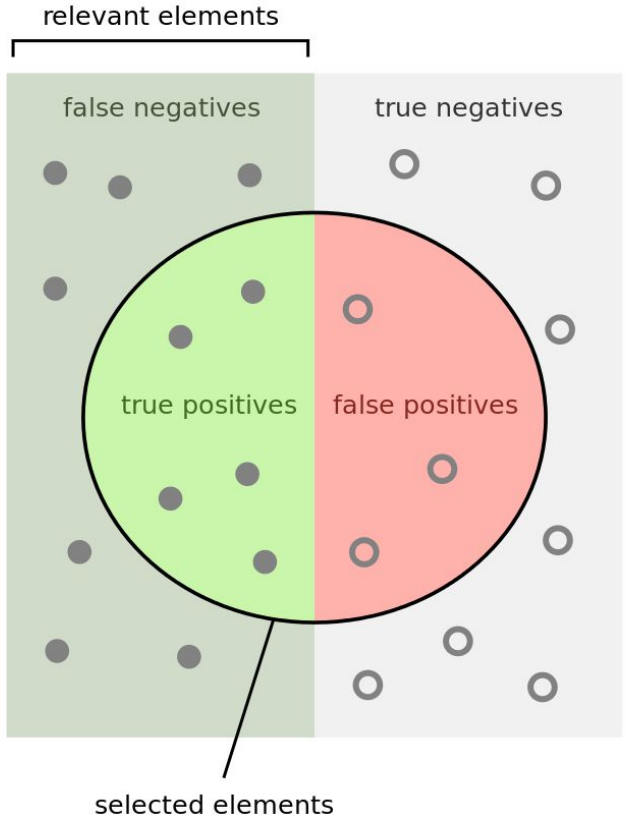
Точность

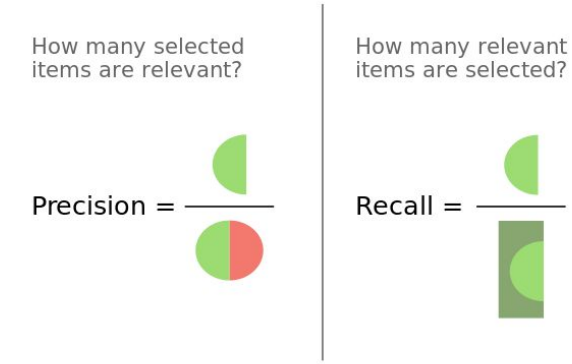
$$Accuracy = \frac{correct}{total}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 = \frac{2}{\frac{1}{precision} + \frac{1}{recall}} = \frac{2 * (precision * recall)}{precision + recall}$$





[Wikipedia]

Как сравнивать?

Бинарная классификация
Binary classification

v
s

Точность

$$Accuracy = \frac{correct}{total}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

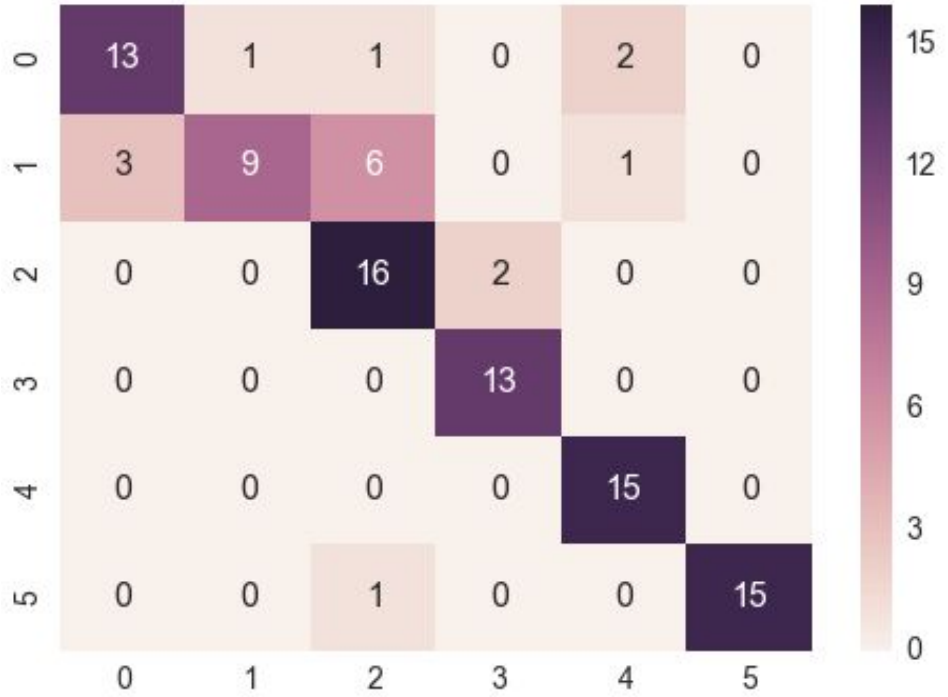
$$F_1 = 2 * \frac{precision * recall}{precision + recall}$$

Как сравнивать?

Многоклассовая классификация
Multi-class classification

Точность

$$Accuracy = \frac{correct}{total}$$



$$Precision_c = \frac{A_{c,c}}{\sum_{i=1}^{n} A_{c,i}}$$

$$Recall_c = \frac{A_{c,c}}{\sum_{i=1}^{n} A_{i,c}}$$

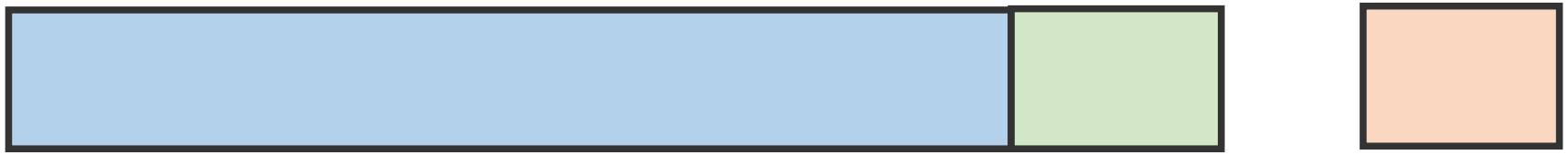
$$Precision = \frac{\sum_{c=1}^{n} P_c}{n}$$

$$Recall = \frac{\sum_{c=1}^{n} R_c}{n}$$

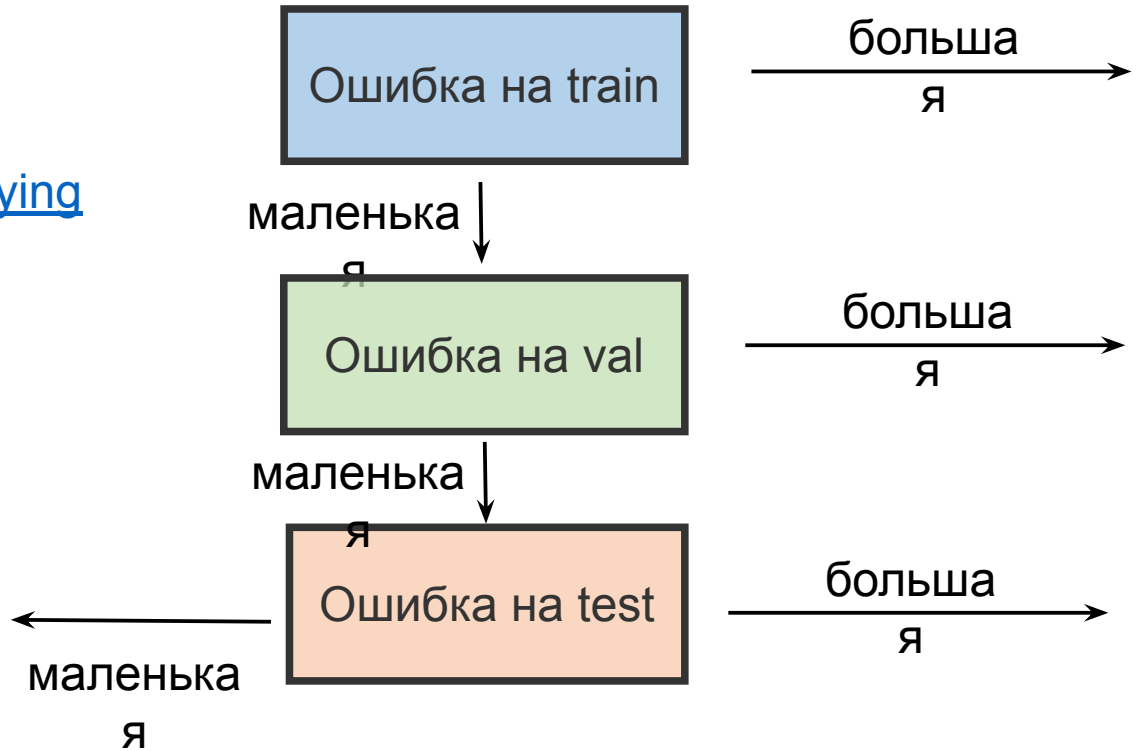
Stack


MAKE GIFS AT GIFSOUP.COM

Machine Learning Flow



Кстати вот он про это подробнее:
[Nuts and Bolts of Applying Deep Learning]

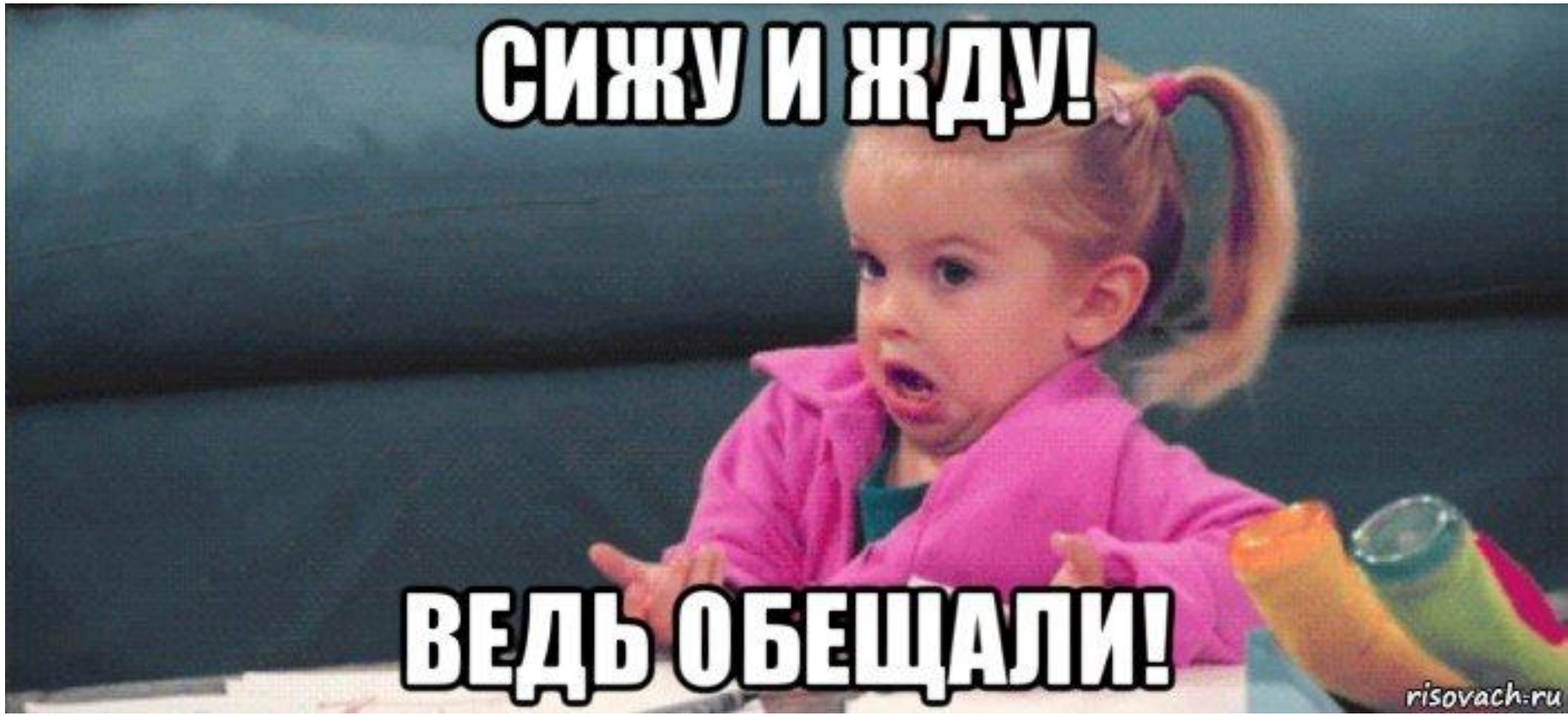


- Более мощную модель
- Больше ресурсов для тренировки
- Другой подход

- Больше данных
- Больше регуляризации
- Другой подход

- Отличаются train и test
- Больше данных, таких как test

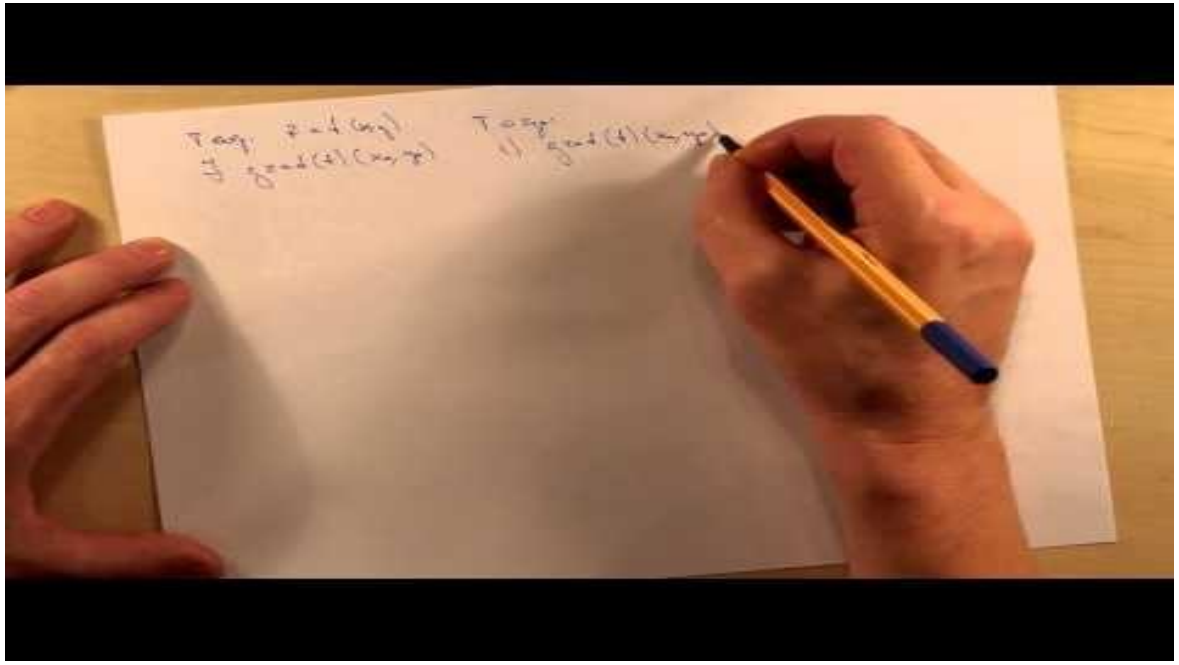
В следующий раз уже будет про нейросети



Домашнее задание!

Повторить производную сложной функции (chain rule)

Link



Link